# SOFTWARE ENGINEERING MEETS CANCER RESEARCH:

**ENABLING INTEROPERABILITY OF DATA AND SERVICES** Anthony Finkelstein<sup>1</sup>, Jeff Kramer<sup>2</sup>, Helen Parkinson<sup>4</sup>, Vito Perrone<sup>1</sup> Fiona Reddington<sup>3</sup>

<sup>1</sup>Department of Computer science, UCL artment of Computer Science, Imperial College <sup>3</sup>NCRI Informatics Coordination Unit <sup>3</sup>European Bioinformatics Institute



## **Project Context**

NCRI Informatics Initiative Aim

"The function of the Informatics Platform is to present resources related to cancer information as a coordinated and collated asset, allowing more efficient access to cancer data and services

Proiect Aim The Platform Reference Model (PRM) project is working collaboratively with the NCRI Informatics Initiative to

establish and document requirements for the platform and to construct and validate the key information models around which the platform will be built

## **Project Objectives**

- · Define the system scope and Identify user and architectural requirements for the NCRI Platform
- Identify integration needs and understand the role different existing resources could play
- Define a Reference Model for the Platform to describe existing data and service resources
- Validate requirements and integration approach by way of use cases and a prototype system

## **Analysis Process**



### Example of use case analysis (use case originally presented to a previous caBIG meting)



## **Reference Model Architecture**

NERI



London

## Top-Down View on Interoperability – An Example

- Existing resources are described by way of the common Platform Reference Model concepts
  Common concepts provide coarse-grain semantics to describe data stored in the repositories
  - · Each dimension describes the resource from a different perspective

### . In this example two application schemas are described from the Cancer Biology perspective



· Once the coarse-grain semantics has been identified, Common Data Elements are used to describe the fine-grain semantics

· Each Object in the Reference Model contains a number of attributes representing CDEs that may be used to describe the fine-grain semantics

• The Reference Model should provide guidance but should not force any choice • Using CDEs stored in caDSR will enable full interoperability with the caBIG network



## Examples of how the model has been used

• The model is embedded in our Prototype User Interface (CRESS)

- Data contained in different repositories can be easily compared
- Semantic Interoperability can be enhanced





## **Conclusions and Next Steps**

### The Analysis Approach

- Systematic user centered way to represent research investigations so that interoperability requirements can be clearly elicited
- Effective in supporting communication in multidisciplinary teams
- **Domain Model and Interoperability Approach**
- Multidimensional domain model to tackle scalability issues
- Top-Down approach to look at different resources in a coherent way
- Fully interoperable and compatible with caBIG

### **Next Steps**

- Understand how the Reference Model can be created and managed (Cooperation with the Large Scale Model Harmonization group in caBIG)
- · Develop a prototype architecture to validate requirements and approach
- · Enable the creation of a unique global community by cooperating with caBIG

References about the Platform Reference Model project • Vito Perone, Anthony Finkelstein, Leah Coldin, Jeff Kramer, Helen Parkinson, Fiona Reddington, "Developing an Integrative Platform for Cancer Research: a Requirements Engineering Perspective". Fifth e-science All Hands Meeting, 18th - 21st September 2006, Nottingham, UK + PRM Project's Web site: <u>www.csu.ca.uk.ca.uk.cancerInformatics</u> NCRI Informatics Initiative's Web site: www.cancerinformatics.org.uk